Biodiversity Data Journal 7: e46833
doi: 10.3897/BDJ.7.e46833

open access

Research Article

From marine park to future genomic observatory? Enhancing marine biodiversity assessments using a biocode approach

Yin Cheong Aden Ip[‡], Ywee Chieh Tay[§,|], Su Xuan Gan[‡], Hui Ping Ang[¶], Karenne Tun[¶], Loke Ming Chou[‡,#], Danwei Huang[‡,#], Rudolf Meier[‡,#]

‡ Department of Biological Sciences, National University of Singapore, Singapore, Singapore
§ National University of Singapore, Singapore, Singapore
| Temasek Life Sciences Laboratory, Singapore, Singapore
¶ National Parks Board, Singapore, Singapore
# Tropical Marine Science Institute, National University of Singapore, Singapore, Singapore

Corresponding author: Ywee Chieh Tay (yweechieh@gmail.com)

Academic editor: Anne Thessen

Received: 25 Sep 2019 | Accepted: 21 Nov 2019 | Published: 10 Dec 2019

Citation: Ip YCA, Tay YC, Gan SX, Ang HP, Tun K, Chou LM, Huang D, Meier R (2019) From marine park to future genomic observatory? Enhancing marine biodiversity assessments using a biocode approach. Biodiversity Data Journal 7: e46833. https://doi.org/10.3897/BDJ.7.e46833

Abstract

Few tropical marine sites have been thoroughly characterised for their animal species, even though they constitute the largest proportion of multicellular diversity. A number of focused biodiversity sampling programmes have amassed immense collections to address this shortfall, but obstacles remain due to the lack of identification tools and large proportion of undescribed species globally. These problems can be partially addressed with DNA barcodes ("biocodes"), which have the potential to facilitate the estimation of species diversity and identify animals to named species via barcode databases. Here, we present the first results of what is intended to be a sustained, systematic study of the marine fauna of Singapore's first marine park, reporting more than 365 animal species, determined based on DNA barcodes and/or morphology represented by 931 specimens (367 zooplankton, 564 macrofauna including 36 fish). Due to the lack of morphological and molecular identification tools, only a small proportion could be identified to species solely based on either morphology (24.5%) or barcodes (24.6%). Estimation of species numbers

© Ip Y et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

for some taxa was difficult because of the lack of sufficiently clear barcoding gaps. The specimens were imaged and added to "Biodiversity of Singapore" (http://singapore.biodiversity.online), which now contains images for > 13,000 species occurring in the country.

Keywords

DNA barcoding, marine park, genomic observatory, COI, biocodes

Introduction

In recent decades, it has become clear that biodiversity loss is an increasingly serious problem and many species are expected to become extinct before discovery and description (Costello et al. 2013, Laurance 2013). It is thought that only 226,000 of the estimated 0.7–1 million marine species have been described (Appeltans et al. 2012). Poor sampling of marine fauna in biodiverse regions and a large backlog of species that have yet to be described have rendered most marine species unidentifiable and often unknown to science (Mora et al. 2011, Mora et al. 2013). This incomplete knowledge of species diversity prevents accurate biodiversity assessments and monitoring and limits our understanding of ecosystem functioning (Isaac et al. 2004). Determining species diversity, using traditional taxonomic techniques, requires skilled taxonomists to accurately identify or describe species based on detailed keys and careful study of morphology. However, this approach is manpower-intensive, slow (Miller 2007) and costly (approximately $39,000–122,000 per species; Carbayo and Marques 2011). Consequently, alternative strategies are being developed to expedite the processes of species discovery and delimitation (Wang et al. 2018). While convincing solutions for large-scale species description are lacking, the problem is starting to attract the attention of many animal taxonomists (Riedel et al. 2013).

Molecular techniques have dramatically increased the rate of species discovery and facilitated species identification for those species that have been barcoded. DNA barcoding was initially proposed as a means to identify animal species, although it is now increasingly used for species discovery (Hebert et al. 2003, Ratnasingham and Hebert 2007, Hajibabaei et al. 2007, Goldstein and DeSalle 2011). This technique uses a short DNA fragment as a standard marker for species description and discovery. For metazoans, the mitochondrial cytochrome c oxidase subunit I (COI) is the barcoding locus of choice, having been popularised by Hebert et al. 2003. However, as a prerequisite to successful species identification via barcodes, a comprehensively curated reference database is required (Ekrem et al. 2007). Public databases, such as GenBank (Benson et al. 2018) and the Barcode of Life Data (BOLD) System (Ratnasingham and Hebert 2007), collect such reference sequences and, while they contain many sequences for vertebrates (Hebert et al. 2004, Ward et al. 2005, Kerr et al. 2007), the coverage for invertebrate species is more limited (Barrett and Hebert 2005, Hajibabaei et al. 2006, Elias et al. 2007, Grant and Linse 2009, Hausmann et al. 2011, Park et al. 2011). These databases are continually updated

as new discoveries are made and presently the BOLD System contains > 5 million barcode sequences belonging to 262,679 species (as of December 2017). However, these identifications should be interpreted with caution, because many are for predicted species (Barcode Index Numbers; "BINS") and few specimens have been formally identified or verified as species (Kwong et al. 2012). Despite these potential issues, research is uncovering unexpected diversity for many taxa in many habitats (Hebert et al. 2004, Brower 2006, Witt et al. 2006, Havermans et al. 2011). These discoveries further reinforce the utility of DNA barcoding in surveys of biodiversity (Smith et al. 2005).

For highly biodiverse regions such as Southeast Asia, these global reference databases remain particularly incomplete and poorly curated (Giam et al. 2010, Jinbo et al. 2011), partially due to the prohibitive costs associated with molecular sequencing (Meier 2008). The sheer number of species found in biodiversity hotspots also poses a considerable challenge, as many of the barcodes recovered differ from those in the databases by more than 3%, meaning accurate species identification is not possible for these animals (Kwong et al. 2012) and several of which may also be new species (Wong et al. 2011, Wang et al. 2018). These problems will only be resolved by greater sampling effort and expansion of curated databases. Unfortunately, a shortage of taxonomic expertise for biodiverse regions is compounded by a lack of the necessary skills required to perform the field collections, a consequence of the decline in biodiversity appreciation (Ríos-Saldaña et al. 2018). It is here that digital reference collections, such as "Biodiversity of Singapore" (https://singapore.biodiversity.online/), can make a difference by helping to stimulate an interest in biodiversity and conservation, with verifying putative species identifications (Will et al. 2005, Ang et al. 2013, Chan et al. 2014).

Singapore is situated just outside the southwest corner of the biodiverse Coral Triangle biodiversity hotspot and, like other countries in the region, its marine biodiversity remains relatively poorly understood. To address this shortfall, we describe the first results of a programme that aims to build a comprehensive animal species identification database for Singapore's first marine park—the Sisters' Islands Marine Park (SIMP; Fig. 1). Prior to this study, preliminary estimates suggested that the SIMP may be home to > 100 fish and > 1,000 macroinvertebrate species (K. Tun, pers. obs.). Barcoding the fauna is part of a larger initiative to make Singapore's biodiversity identifiable with molecular tools and constitutes the first steps towards building a national genomic observatory (Davies et al. 2014). Located approximately 6 km south of mainland Singapore, SIMP is recognised as a locally important area of biodiversity in terms of coral species richness, functional and phylogenetic diversity (Wong et al. 2018). Previous work modelling coral larval dispersal has indicated that reefs in the area are potentially strong source reefs that can seed other reefs in Singapore (Tay et al. 2012, Chang 2015). Furthermore, the marine park also aims to serve as an outreach platform to encourage public interest in marine life and participation in biodiversity conservation. It is open to the public and visitors can interact with the natural environment via guided intertidal walks and subtidal dive trails. Thus, a systematic and regular documentation of the biodiversity at the SIMP is important for the management and conservation of marine ecosystems in Singapore. SIMP was established as a marine park on 15 July 2015 and spans an area of 40 hectares that encompasses the